Eur. Phys. J. Spec. Top.
https://doi.org/10.1140/epjs/s11734-025-02011-1

THE EUROPEAN PHYSICAL JOURNAL SPECIAL TOPICS

Regular Article

A topological graph kernel based on Q-analysis of clique complexes

Nikita Smirnov[1,a], Semen Kurkin[2,b], and Alexander Hramov[2,c]

[1] Baltic Center of Neurotechnology and Artificial Intelligence, Immanuel Kant Baltic Federal University, Alexandra Nevskogo 14, 236041 Kaliningrad, Russia
[2] Research Institute of Applied Artificial Intelligence and Digital Solutions, Plekhanov Russian University of Economics, 36 Stremyannyy Ln., 115093 Moscow, Russia

Received 6 August 2025 / Accepted 1 October 2025

Abstract Graph classification methods struggle to capture higher-order network structures beyond pairwise links, as many approaches are bounded by the 1-dimensional Weisfeiler-Lehman test and cannot distinguish graphs with similar local neighborhoods but different global topologies. We propose the Q-Analysis Kernel (QAK), which follows a systematic pipeline: input graphs are transformed into maximal clique simplicial complexes using the Bron–Kerbosch algorithm, then Q-analysis extracts topological feature vectors including the First Structure Vector (FSV), Third Structure Vector (TSV), and Topological Entropy. These vectors quantify the multi-scale connectivity of higher-order structures and are used with a cosine kernel for SVM classification. Experimental evaluation on TUDataset benchmarks demonstrates that QAK achieves classification performance comparable to established graph kernels, with competitive accuracies on social network datasets (73.7% on IMDB-BINARY, 75.4% on REDDIT-BINARY) compared to Weisfeiler–Lehman and Shortest Path kernels. The method provides a novel, interpretable graph representation rooted in algebraic topology, where each feature component corresponds to quantifiable topological properties. This enables feature importance analysis to reveal which structural scales (q-levels) are most discriminative for classification decisions.

1 Introduction

The study of complex systems—from biological networks to engineered infrastructures—demands tools capable of capturing not only pairwise interactions but also the higher-order structures that define their emergent behavior. While Graph Neural Networks (GNNs) and graph kernels have become cornerstone methods for graph classification, their expressive power is fundamentally limited by the 1-dimensional Weisfeiler-Lehman (1-WL) test [14, 16]. This constraint renders them unable to distinguish between graphs with identical local neighborhoods but divergent global topologies, such as a 6-cycle versus two disconnected 3-cycles [8]. This classic limitation is illustrated in Fig. 1. While the 1-WL test cannot distinguish Graph A from Graph B, their topological signatures derived from Q-analysis are clearly different. Such limitations are particularly consequential in systems where functional properties arise from multi-node interactions, including molecular activity in cheminformatics [9] or mesoscale connectivity patterns in neural networks [3].

Recent work underscores the critical role of higher-order interactions in functional networks, particularly in neuroscience. For instance, *Q-analysis* of fMRI-based brain networks in major depressive disorder (MDD) revealed disrupted higher-order topology, including reduced clique connectivity and altered limbic-cerebellar integration [11]. Similarly, consensus network approaches have identified segregation deficits in MDD, with the clustering coefficient emerging as a high-precision biomarker [10]. These findings highlight how traditional pairwise metrics may overlook clinically relevant mesoscale structures. Even advanced tools like GNNs face challenges: while they achieve high accuracy in MDD classification (93%), their performance depends critically on preprocessing steps like graph

[a] e-mail: nikita.smirnov.m@gmail.com (corresponding author)
[b] e-mail: kurkinsa@gmail.com
[c] e-mail: hramov.ae@rea.ru

Published online: 18 October 2025

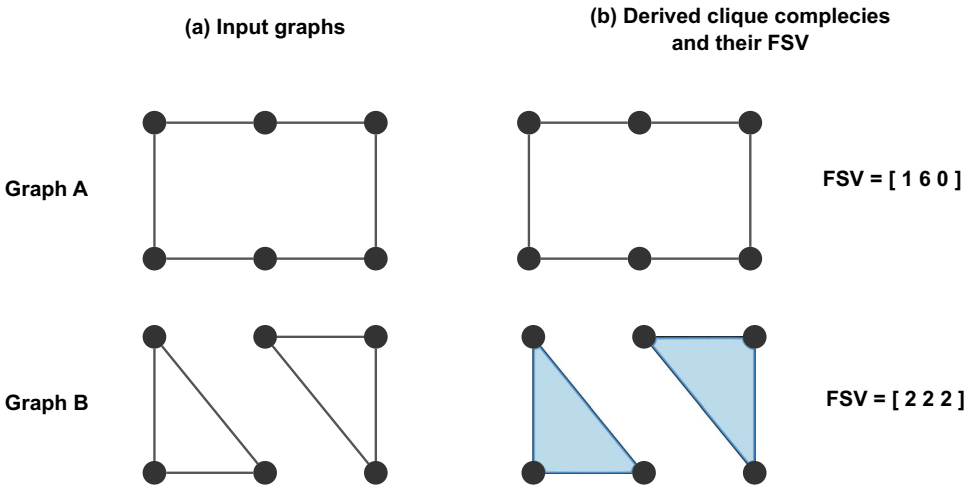

Fig. 1 An illustration of the 1-WL test's limitation and the discriminative power of Q-analysis. **a** Two non-isomorphic graphs: a 6-cycle (Graph A) and two disconnected 3-cycles (Graph B). Methods bounded by the 1-WL test cannot distinguish them. **b** The derived clique complexes and their First Structure Vectors (FSV). Graph A contains only 1-simplices (edges) and is fully connected, while Graph B contains two disconnected 2-simplices (triangles). This results in distinct FSVs, demonstrating that topological signatures can capture global structural differences that local methods miss. The FSV is shown as $[Q_0, Q_1, Q_2]$

sparsification to avoid overfitting [18]. Hypergraph representations, conversely, enhance detection of conditions like autism by explicitly modeling multilayer brain interactions [17], demonstrating the value of topology-aware methods.

To address these gaps, we propose a framework leveraging the *clique complex*–a simplicial complex constructed from maximal cliques [2]–and *Q-analysis* [1]. Our contribution is a novel graph kernel, the Q-Analysis Kernel (QAK), derived from these topological signatures. The empirical results show that this kernel achieves performance comparable to established methods like the Shortest Path kernel, while not outperforming more powerful kernels based on the Weisfeiler-Lehman hierarchy. Therefore, the primary contribution of this work is a methodological framework for generating interpretable, topology-based graph representations.

Our approach not only circumvents the 1-WL bottleneck but also opens avenues for analyzing systems governed by nonlocal interactions, a critical challenge in the study of far-from-equilibrium physical and biological networks.

The hypothesis is that these Q-analytic signatures provide a meaningful representation of a graph's higher-order structure, and that this representation is sufficiently discriminative to be a viable tool for graph analysis. The key benefits of this approach are:

- Interpretability: The features are not learned embeddings, but are quantifiable measures of underlying clique complex topology. This allows for post-hoc analysis to understand precisely which structural properties are driving classification decisions.
- A Complementary Structural Lens: The QAK analyzes graph structure through the global connectivity of cliques, a perspective fundamentally different from the local neighborhood aggregation of GNNs or the shortest-path distances used in other kernels. This makes it a valuable tool for comparative structural analysis.

By introducing and evaluating this kernel, we demonstrate that Q-analysis provides a sound basis for building graph representations that, while not necessarily superior in predictive power, offer a unique and transparent view into the higher-order organization of complex networks.

2 Background

2.1 From graphs to simplicial complexes

A *graph* is a pair $G = (V, E)$, where V is a set of vertices and $E \subseteq \binom{V}{2}$ is a set of edges connecting pairs of vertices.

A *clique* in a graph G is a subset of vertices $C \subseteq V$ where every two distinct vertices in C are adjacent. A clique C is maximal if it is not a subset of any larger clique.

An abstract simplicial complex K on a set of vertices V is a collection of subsets of V, called *simplices*, such that if $\sigma \in K$ and $\tau \subset \sigma$, then $\tau \in K$. The dimension of a simplex σ is defined as $\dim(\sigma) = |\sigma| - 1$. A 0-simplex is a vertex, a 1-simplex is an edge, a 2-simplex is a triangle, and so on.

The *clique complex* of a graph G, denoted $K(G)$, is a simplicial complex whose simplices are the cliques of G. To construct $K(G)$, we identify the maximal cliques of G. These maximal cliques, along with all of their subsets (which are also cliques), form the set of all simplices in the complex.

2.2 Q-analysis

Q-analysis is a method from algebraic topology designed to study the global connectivity structure of a simplicial complex [1]. It analyzes how simplices are connected to one another at different dimensional levels. The process is illustrated in Fig. 2, which shows how a graph (a) is lifted to its clique complex (b) and then analyzed for connectivity between its simplices (c).

The core concepts are built upon a relation of nearness between simplices.

- q-Nearness: Two simplices, σ_i and σ_j, are said to be *q-near* if they share a common face of at least dimension q. Mathematically, this condition is $\dim(\sigma_i \cap \sigma_j) \geq q$. In a clique complex, this is equivalent to sharing at least $q + 1$ common vertices.
- q-Connectivity: The q-nearness relation is extended by transitivity. Two simplices are *q-connected* if there exists a chain of simplices connecting them where each adjacent pair in the chain is at least q-near. This defines an equivalence relation on the set of simplices of dimension $\geq q$.
- q-Connected Components: The equivalence classes defined by the q-connectivity relation are the *q–connected components*. The number of such components at a given dimensional level q is denoted by Q_q.

This framework yields several quantitative metrics that can be used as feature vectors for a graph.

First Structure Vector (FSV) The FSV is an integer vector that captures the number of connected components at each dimensional level. For a complex of maximum dimension d, the FSV is:

$$\mathbf{Q} = (Q_d, Q_{d-1}, \ldots, Q_1, Q_0) \tag{1}$$

where Q_q is the number of q-connected components. The FSV provides a multi-scale signature of the complex's fragmentation.

Third Structure Vector (TSV) The TSV is a normalized version of the FSV that measures the degree of connectivity relative to the number of available simplices. Let n_q be the number of simplices with dimension

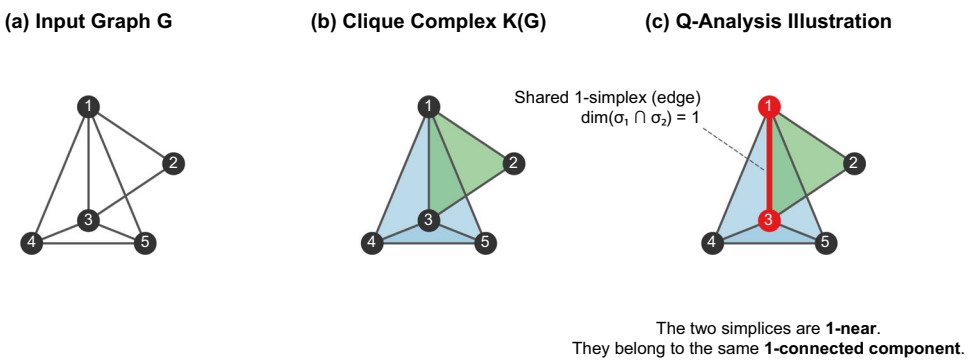

Fig. 2 Visual explanation of the Q-analysis framework. **a** An input graph G. **b** The corresponding clique complex K(G) is constructed. It contains two maximal 2-simplices (the filled triangles 1, 2, 3 and 1, 3, 5). **c** An illustration of Q-analysis concepts. The two 2-simplices are 1-near because they share a 1-simplex (the edge between vertices 1 and 3). Because they are 1-near, they belong to the same 1-connected component

greater than or equal to q. The TSV is a vector $\hat{\mathbf{Q}}$ with components:

$$\hat{Q}_q = 1 - \frac{Q_q}{n_q} \tag{2}$$

A value of $\hat{Q}_q \approx 1$ indicates high connectivity (many simplices form few components), while $\hat{Q}_q = 0$ indicates total fragmentation ($Q_q = n_q$, where each simplex is its own component).

Topological Entropy Topological entropy quantifies the diversity of vertex participation in simplices at each dimensional level q. Let $Q_i^{(q)}$ be the number of q-dimensional simplices containing vertex i, and let $p_i^{(q)} = Q_i^{(q)} / \sum_j Q_j^{(q)}$ be the normalized participation of vertex i. The topological entropy, normalized to [0, 1], is:

$$S_Q(q) = -\frac{\sum_i p_i^{(q)} \log p_i^{(q)}}{\log M_q} \tag{3}$$

where M_q is the number of vertices that participate in at least one q-simplex. High entropy indicates that vertices participate evenly in the higher-order structures, while low entropy suggests that a few central vertices dominate these structures.

3 Methodology: the Q-analysis kernel (QAK)

The QAK is constructed through a two-stage process. First, each graph is transformed into a fixed-length feature vector via a topological embedding pipeline. Second, these vectors are used to compute a kernel matrix for a Support Vector Machine classifier.

3.1 Graph embedding pipeline

The pipeline converts a graph G into a feature vector $\mathbf{v}_G$ that encodes its higher-order connectivity. The process consists of the following steps:

Step 1: *Input* The process takes a graph $G = (V, E)$ as input.
Step 2: *Clique Complex Construction* The graph is lifted into a higher-dimensional representation by constructing its clique complex, $K(G)$. This is achieved by first identifying the set of all maximal cliques in G using the Bron-Kerbosch algorithm [5]. These maximal cliques form the simplices of the complex $K(G)$.
Step 3: *Q-Analysis Feature Extraction* The Q-analysis metrics defined in Sect. 2.2 are computed for the complex $K(G)$. Specifically, we calculate the First Structure Vector ($\mathbf{Q}$), the Third Structure Vector ($\hat{\mathbf{Q}}$), and the Topological Entropy vector ($\mathbf{S}_Q$). These vectors can be used individually or concatenated to form a comprehensive feature vector for the graph. For instance, a combined feature vector might be $\mathbf{v}_G = [\mathbf{Q} \parallel \hat{\mathbf{Q}} \parallel \mathbf{S}_Q]$, where $\parallel$ denotes concatenation.
Step 4: *Vector Standardization* The maximum dimension of clique complexes can vary between graphs, leading to Q-analysis vectors of different lengths. To create a consistent feature space for the kernel, we standardize the vector lengths. We first determine the maximum simplex dimension, $d_{\max}$, observed across all graphs in the training dataset. Any feature vector derived from a graph with a maximum dimension less than $d_{\max}$ is padded with zeros to match the length corresponding to a $d_{\max}$-dimensional complex.
Step 5: *Output* The final output of the pipeline is a fixed-length feature vector $\mathbf{v}_G$ that serves as the graph's topological signature.

3.2 Kernel computation and classification

Once each graph G_i in a dataset is represented by its feature vector $\mathbf{v}_{G_i}$, we can define the kernel. The similarity between two graphs is measured using the cosine kernel. This choice is motivated by the fact that Q-analysis vectors can vary significantly in magnitude depending on the size and complexity of the underlying graph. The cosine kernel normalizes for magnitude, focusing the comparison on the orientation, or the relative profile, of the feature vectors, which is well-suited for comparing topological signatures across graphs of different scales. The

Table 1 Statistics for the TUDataset benchmarks used in the evaluation. All data are sourced from Morris et al. [15]

Dataset	# Graphs	# Classes	Avg. Nodes	Avg. Edges
ENZYMES	600	6	32.6	62.1
IMDB-BINARY	1,000	2	19.8	96.5
IMDB-MULTI	1,500	3	13.0	65.9
NCI1	4,110	2	29.9	32.3
PROTEINS	1,113	2	39.1	72.8
REDDIT-BINARY	2,000	2	429.6	497.8
MCF-7	27,770	2	26.4	28.5
MOLT-4	39,765	2	26.1	28.1
YEAST	79,601	2	21.5	22.8
GITHUB_STAR	12,725	2	113.8	234.6
REDDIT_THREADS	203,088	2	23.9	25.0

kernel function is defined as:

$$k(G_i, G_j) = \frac{\mathbf{v}_{G_i} \cdot \mathbf{v}_{G_j}}{\|\mathbf{v}_{G_i}\| \|\mathbf{v}_{G_j}\|} \tag{4}$$

The resulting Gram matrix, containing the pairwise kernel evaluations for all graphs in the dataset, is then supplied to a standard Support Vector Machine (SVM) for training and classification [7]. While not the focus of this work, it is worth noting that this feature representation is inherently interpretable. Each component of the vector $\mathbf{v}_G$ corresponds to a specific topological property, making it possible to conduct post-hoc feature importance analysis to understand which structural characteristics are most discriminative for a given task.

4 Experimental evaluation

This section details the experimental setup used to evaluate the performance of the QAK. To provide a meaningful benchmark, we compare our results against established graph kernels on a standard collection of datasets.

4.1 Datasets

All datasets used in this study are from the **TUDataset** [15]. We selected a range of small and mid-scale datasets from bioinformatics and social network domains to evaluate our method across different structural properties. A summary of the datasets is provided in Table 1. All node and edge labels were disregarded to focus the evaluation purely on graph structure, consistent with the protocol for the baseline methods.

4.2 Baseline methods

The performance of our proposed QAK is compared against a set of standard graph kernels. The accuracy scores for these baseline methods are taken directly from the comprehensive experimental study presented in Morris et al. [15]. The selected baselines are:

- **Weisfeiler-Lehman Subtree Kernel (1-WL)** [6].
- **Graphlet Kernel (GR)** [19].
- **Shortest-Path Kernel (SP)** [4].

4.3 Experimental protocol

To ensure a fair and direct comparison, we precisely followed the experimental protocol established by Morris et al. [15] for their evaluation of graph kernels.

The classification performance is measured by **mean classification accuracy** with standard deviation. The evaluation is performed using a **10-fold cross-validation** scheme. To ensure the robustness of the results, this entire 10-fold procedure is repeated 10 times with different random seeds for data splitting. The final accuracy is the average over all 100 runs.

For all kernel methods, including our own, a **C-Support Vector Machine (C-SVM)** is used as the classifier. The regularization parameter C of the SVM is optimized for each fold via a search over the set $\{10^{-3}, 10^{-2}, \ldots, 10^{2}, 10^{3}\}$.

Our Q-Analysis Kernel (QAK) was evaluated using the same protocol. We tested several variants of the kernel based on different combinations of the Q-analysis feature vectors:

- **QAK-FSV:** Kernel based on the First Structure Vector.
- **QAK-TSV:** Kernel based on the Third Structure Vector.
- **QAK-ENT:** Kernel based on the Topological Entropy vector.
- **QAK-FSV-ENT:** Kernel based on the concatenation of FSV and Entropy vectors.
- **QAK-TSV-ENT:** Kernel based on the concatenation of TSV and Entropy vectors.

For each variant, the corresponding feature vectors were computed for all graphs and then used to build a Gram matrix with the cosine kernel, which was subsequently fed into the C-SVM classifier following the 10-fold cross-validation procedure.

5 Results and discussion

This section presents and analyzes the classification performance of the Q-Analysis Kernel (QAK) variants in comparison to established baseline methods. We also discuss the inherent limitations of the proposed approach.

5.1 Classification performance

The classification accuracies of the QAK variants and baseline kernels on both small- and mid-scale datasets are summarized in Tables 2 and 3. Baseline results are taken from Morris et al. [15], while QAK results were obtained using the experimental protocol described in Sect. 4.3.

Analysis of results The performance of the Q-Analysis Kernel varies significantly across datasets and feature combinations. On social network datasets like **IMDB-BINARY** and **IMDB-MULTI**, the QAK-TSV-ENT variant performs competitively, achieving an accuracy of 73.7% on IMDB-BINARY, which is on par with the 1-WL kernel (72.4%). On **REDDIT-BINARY**, the QAK-FSV-ENT variant (75.4%) underperforms the Shortest-Path kernel (84.5%) but is slightly better than the 1-WL kernel (73.3%). Similarly, on **REDDIT_THREADS**, the QAK-FSV-ENT variant (74.8%) is competitive with the SP kernel (77.3%). This suggests that the connectivity of cliques is a discriminative feature in certain social network graphs.

Table 2 Classification accuracies (%) on small-scale TUDatasets. Baselines from Morris et al. [15]

Method	ENZYMES	IMDB-B.	IMDB-M.	NCI1	PROTEINS	REDDIT-B.
Baselines						
1-WL	50.8 ± 7.1	72.4 ± 4.4	50.5 ± 3.6	84.2 ± 1.8	72.6 ± 3.4	73.3 ± 3.0
GR	29.5 ± 5.4	59.8 ± 4.9	39.5 ± 4.0	66.0 ± 2.4	71.6 ± 4.0	59.7 ± 3.8
SP	39.3 ± 7.1	58.4 ± 5.3	39.4 ± 4.4	74.2 ± 2.1	75.6 ± 4.0	84.5 ± 2.5
Q-Analysis Kernels (QAK)						
QAK-FSV	21.6 ± 5.2	72.2 ± 4.3	49.6 ± 3.9	52.4 ± 3.0	71.0 ± 4.0	75.3 ± 3.1
QAK-TSV	18.8 ± 4.5	64.5 ± 4.8	40.2 ± 3.6	48.6 ± 1.9	62.7 ± 4.8	54.2 ± 4.3
QAK-ENT	21.1 ± 5.7	70.3 ± 4.5	48.5 ± 4.2	51.0 ± 2.4	67.7 ± 4.6	55.5 ± 3.5
QAK-FSV-ENT	23.2 ± 5.0	72.9 ± 4.2	49.4 ± 3.9	**64.6 ±2.0**	**71.8 ±4.0**	**75.4 ±2.8**
QAK-TSV-ENT	**23.7 ±5.5**	**73.7 ±3.8**	**50.7 ±3.8**	50.8 ± 2.9	67.7 ± 4.2	69.0 ± 3.3

The bold values highlight the best classification accuracy achieved by a Q-Analysis Kernels (QAK) variant for each specific dataset

Table 3 Classification accuracies (%) on mid-scale TUDatasets

Method	MCF-7	MOLT-4	YEAST	GITHUB_STAR	REDDIT_THR.
Baselines					
1-WL	94.5 ± 0.3	94.6 ± 0.4	89.2 ± 0.4	64.0 ± 1.4	77.0 ± 0.3
GR	91.7 ± 0.5	92.1 ± 0.4	88.2 ± 0.3	53.6 ± 1.6	51.2 ± 0.3
SP	91.7 ± 0.6	92.1 ± 0.4	88.2 ± 0.4	64.2 ± 1.3	77.3 ± 0.2
Q-Analysis Kernels (QAK)					
QAK-FSV	91.7 ± 0.4	92.1 ± 0.3	88.0 ± 0.4	56.6 ± 1.3	63.6 ± 0.3
QAK-TSV	91.7 ± 0.5	92.1 ± 0.4	88.0 ± 0.3	56.7 ± 2.1	51.2 ± 0.4
QAK-ENT	91.7 ± 0.5	92.1 ± 0.5	88.0 ± 0.2	**61.4 ±1.4**	62.5 ± 0.4
QAK-FSV-ENT	91.7 ± 0.5	92.1 ± 0.4	88.0 ± 0.3	57.0 ± 1.2	**74.8 ±0.3**
QAK-TSV-ENT	**91.7 ±0.5**	**92.1 ±0.4**	**88.0 ±0.3**	61.0 ± 0.9	70.4 ± 0.4

The bold values highlight the best classification accuracy achieved by a Q-Analysis Kernels (QAK) variant for each specific dataset

In contrast, on bioinformatics datasets such as **ENZYMES** and **NCI1**, the QAK variants consistently underperform all baselines. For example, the best QAK result on NCI1 is 64.6%, significantly lower than the 1-WL kernel's 84.2%. This may indicate that the class-distinguishing information in these graphs is contained in local substructures or specific node attributes, which our structure-only method does not capture. On the larger bioinformatics datasets (MCF-7, MOLT-4, YEAST), the QAK performance is respectable but does not surpass the baselines.

Regarding the feature combinations, concatenating a structure vector (FSV or TSV) with the Topological Entropy (ENT) vector generally yields the best results. For instance, QAK-TSV-ENT is the top-performing QAK variant on 5 of the 11 datasets. This suggests that combining a measure of component fragmentation (FSV/TSV) with a measure of vertex participation diversity (ENT) creates a more powerful representation than using either metric alone. The normalized connectivity measured by TSV seems particularly effective for social network graphs, while the raw component counts from FSV perform better on the REDDIT datasets.

5.2 Limitations

The proposed methodology has two primary limitations.

Computational Cost of Clique Enumeration The first step in our pipeline is the identification of all maximal cliques. This problem is known to be NP-hard [5]. While algorithms for maximal clique enumeration are efficient for sparse graphs, their performance degrades significantly on large, dense graphs, where the number and size of cliques can grow exponentially. This computational bottleneck currently restricts the practical application of the QAK to graphs of small to moderate size and density, where clique enumeration remains feasible. Future work could explore heuristic or approximate methods, such as algorithms based on "weak cliques", to improve scalability for very large graphs [20].

Exclusion of Node and Edge Attributes The current implementation of the QAK operates purely on the topological structure of the graph, disregarding any attributes associated with nodes or edges. In many real-world applications, such as cheminformatics, these attributes (e.g., atom type, bond order) carry critical information for classification. The baselines, particularly the 1-WL kernel, naturally incorporate discrete node labels into their procedure. The inability of our current method to leverage this information is a significant factor contributing to its underperformance on datasets like NCI1. Integrating node and edge attributes into the Q-analysis framework is a non-trivial but essential direction for future work.

6 Conclusion

In this thesis, we introduced and evaluated the Q-Analysis Kernel, a novel, topology-based kernel for graph classification. The method provides an alternative lens for analyzing graph structure by moving beyond pairwise links and local neighborhoods to focus on the global connectivity of higher-order clique structures. By lifting a

graph to its clique complex and computing its Q-analytic signature, we generate a feature representation that captures the multi-scale organization of its most cohesive subgraphs.

Our experimental evaluation on a range of benchmark datasets demonstrates that the QAK is a viable method for graph classification. While it does not consistently outperform state-of-the-art methods like the Weisfeiler-Lehman kernel, it achieves competitive performance on several social network datasets where the organization of communities and groups is a key characteristic. The results indicate that the representation of higher-order structure provided by Q-analysis is sufficiently discriminative for classification tasks and offers a valuable complement to existing methods.

Perhaps the most significant contribution of this work is the development of an interpretable graph representation. Unlike the abstract embeddings produced by many learning-based models, each feature in a Q-analysis vector corresponds to a precise, quantifiable topological property. This transparency opens the door for a deeper understanding of the structural drivers of classification, allowing researchers to probe which scales of connectivity are most important for distinguishing between classes of graphs.

7 Future work

A key avenue for future work is the integration of node and edge attributes. The current framework could be extended by defining attribute-aware simplices. For instance, for graphs with discrete node labels, a valid simplex might require all its constituent vertices to share the same label. For continuous attributes, simplices could be weighted based on the similarity of their vertex attributes, leading to a weighted simplicial complex. Applying Q-analysis to such attributed topological structures would allow the model to leverage feature information, likely improving performance on datasets where node attributes are critical. For example, the q-nearness relation could be redefined to depend not only on the dimension of shared faces but also on their associated weights, enabling a more nuanced, feature-aware topological analysis.

Finally, the topological framework presented here could be extended to study collective states in complex dynamical networks with higher-order interactions. The analysis of simplicial complexes has proven fruitful for understanding emergent dynamical phenomena such as synchronization and contagion, which are often governed by group interactions rather than just pairwise connections. Recent reviews have highlighted the importance of higher-order network structures in shaping dynamics in biological and social systems [12, 13]. Future research could explore how Q-analysis signatures correlate with the emergence of collective behaviors, potentially providing a bridge between the static topological structure of a system and its dynamic functional states.

Acknowledgements Nikita Smirnov was supported by the Brain Program of the IDEAS Research Center. Semen Kurkin and Alexander Hramov were also supported by the Russian Science Foundation (project 23-71-30010). The authors would like to thank the members of the Baltic Center of Neurotechnology and Artificial Intelligence for their valuable feedback and discussions.

Data Availability The data presented in this study are available on request from Nikita Smirnov.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. R.H. Atkin, Mathematical structure in human affairs (1974)
2. F. Battiston, G. Cencetti, I. Iacopini, et al, Networks beyond pairwise interactions: structure and dynamics. (2020). arXiv [physicssoc-ph] pp 1–92 [physics.soc-ph]
3. A.R. Benson, D.F. Gleich, J. Leskovec, Higher-order organization of complex networks. Science **353**(6295), 163–166 (2016)
4. K.M. Borgwardt, H. Kriegel, Shortest-path kernels on graphs. In: Fifth IEEE International Conference on Data Mining (ICDM'05). IEEE, p 8 pp (2006)
5. C. Bron, J. Kerbosch, Algorithm 457: finding all cliques of an undirected graph. Commun. ACM **16**(9), 575–577 (1973)
6. Computational Biology Research Group, Weisfeiler-lehman graph kernels. J. Mach. Learn. Res. **12**, 2539–2561 (2011)
7. C. Cortes, V. Vapnik, Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)

8. A. Feng, C. You, S. Wang et al., KerGNNs: interpretable graph neural networks with graph kernels. Proc Conf AAAI Artif Intell **36**(6), 6614–6622 (2022)
9. T. Gaudelet, N. Malod-Dognin, N. Pržulj, Higher-order molecular organization as a source of biological function. Bioinformatics **34**(17), i944–i953 (2018)
10. V.S. Khorev, S.A. Kurkin, G. Zlateva et al., Disruptions in segregation mechanisms in fmri-based brain functional network predict the major depressive disorder condition. Chaos Solit. Fract. **188**, 115566 (2024)
11. S.A. Kurkin, N.M. Smirnov, R. Paunova et al., Beyond pairwise interactions: higher-order Q-analysis of fmri-based brain functional networks in patients with major depressive disorder. IEEE Access **12**, 197168–197186 (2024)
12. S. Majhi, M. Perc, D. Ghosh, Dynamics on higher-order networks: a review. J. R. Soc. Interface **19**(188), 20220043 (2022)
13. S. Majhi, S. Ghosh, P.K. Pal et al., Patterns of neuronal synchrony in higher-order networks. Phys. Life Rev. **52**, 144–170 (2025)
14. H. Maron, H. Ben-Hamu, H. Serviansky, et al, Provably powerful graph networks. Neural Inf Process Syst abs/1905.11136 (2019)
15. C. Morris, N.M. Kriege, F. Bause, et al, TUDataset: A collection of benchmark datasets for learning with graphs. (2020). arXiv [csLG] [cs.LG]
16. C. Morris, Y. Lipman, H. Maron et al., Weisfeiler and leman go machine learning: The story so far. J. Mach. Learn. Res. **24**(333), 1–59 (2023)
17. E. Pitsik, S. Kurkin, O. Martynova, et al, Hypergraph representation of multilayer brain network enhances autism spectrum disorder detection. Chaos: Interdiscip. J. Nonlinear Sci. 35(7). (2025)
18. E.N. Pitsik, V.A. Maximenko, S.A. Kurkin et al., The topology of fmri-based networks defines the performance of a graph neural network for the classification of patients with major depressive disorder. Chaos Solit. Fract. **167**, 113041 (2023)
19. N. Shervashidze, S. Vishwanathan, T. Petri et al., Efficient graphlet kernels for large graph comparison. AISTATS **5**, 488–495 (2009)
20. X. Zhang, C. Wang, Y. Su et al., A fast overlapping community detection algorithm based on weak cliques for large-scale networks. IEEE Trans. Comput. Soc. Syst. **4**(4), 218–230 (2017)